

BEST AVAILABLE COPY

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-258907

(43)Date of publication of application : 03.10.1997

(51)Int.Cl.

G06F 3/06

G06F 3/06

G06F 13/10

G11B 19/02

(21)Application number : 08-068405

(71)Applicant : MITSUBISHI ELECTRIC CORP

(22)Date of filing : 25.03.1996

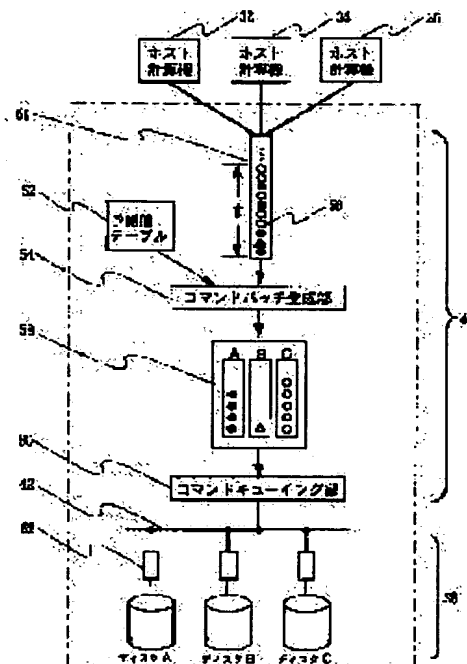
(72)Inventor : NAGASHIMA MASARU

(54) HIGHLY AVAILABLE EXTERNAL STORAGE DEVICE HAVING PLURAL STORAGE DISK PARTS

(57)Abstract:

PROBLEM TO BE SOLVED: To improve a response and throughput ad to eliminate the imbalance of the response and throughput between storage disk parts by most suitably processing the queue of a command issued to an external storage device.

SOLUTION: The command batch generation part 54 of the external storage control part 40 reads the number of commands whose sum of processing time prediction values by a prediction value table 52 becomes prescribed time as a group (batch) from a reception queue 50. The batch is stored in a transmission queue 58. A command queuing part 60 repeats a processing for selecting a disk device 38 whose prediction value of command processing time is the longest, for taking out the command for it and issuing it to the disk device 38. When the transmission queue 58 becomes vacant, the next batch is read into it.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平9-258907

(43) 公開日 平成9年(1997)10月3日

(51) Int.Cl. ⁸	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 3/06	3 0 1		G 0 6 F 3/06	3 0 1 F
	5 4 0			5 4 0
13/10	3 4 0		13/10	3 4 0 B
G 1 1 B 19/02	5 0 1		G 1 1 B 19/02	5 0 1 F

審査請求 未請求 請求項の数10 O L (全 16 頁)

(21) 出願番号 特願平8-68405

(22) 出願日 平成8年(1996)3月25日

(71) 出願人 000006013

三菱電機株式会社

東京都千代田区丸の内二丁目2番3号

(72) 発明者 長島 勝

東京都千代田区丸の内二丁目2番3号 三

菱電機株式会社内

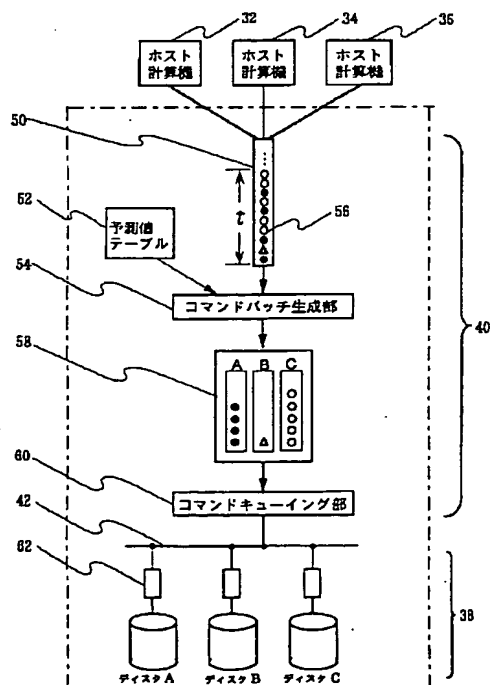
(74) 代理人 弁理士 吉田 研二 (外2名)

(54) 【発明の名称】 複数の記憶ディスク部を有した高可用性の外部記憶装置

(57) 【要約】

【課題】 外部記憶装置に対して発行されたコマンドのキューが最適に処理されておらず、レスポンスやスループットといった性能が不十分である。

【解決手段】 外部記憶制御部40のコマンドバッチ生成部54は、予測値テーブル52によるコマンドの処理時間予測値の和が所定時間となる数のコマンドをグループ(バッチ)として受入キュー50から読み出す。バッチは送出キュー58に格納される。コマンドキューイング部60はその中のコマンド処理時間の予測値が最長のディスク装置38を選択し、それに対するコマンドを取り出して対応するディスク装置38に発行するという動作を繰り返す。送出キュー58は空になると、これに次のバッチが読み込まれる。



【特許請求の範囲】

【請求項1】 データを格納しこのデータに対する読み出し／書き込みコマンドをキューイング可能な複数の記憶ディスク部と、コンピュータからの前記コマンドを受入キューに蓄え前記記憶ディスク部に順次発行し、前記コンピュータと前記記憶ディスク部との間のデータ入出力処理を制御する外部記憶制御部とを備えた外部記憶装置において、

前記外部記憶制御部は、

前記コマンドの処理に要すると予測される時間であるコマンド処理時間予測値を生成する予測処理時間生成手段と、

所定のタイミングで前記受入キュー内のコマンドをその前記コマンド処理時間予測値の和が所定の処理タイムスライスに応じた値となる個数だけ格納される、前記コマンドの待ち行列である送出キューと、

前記予測処理時間生成手段に基づき前記各記憶ディスク部別の処理時間を予測する処理時間予測手段と、

この予測された処理時間が最大の前記記憶ディスク部に対する前記コマンドを前記送出キューから取り出して、対応する前記記憶ディスク部にキューイングするコマンドキューイング手段と、

前記各送出キューが空になると、前記受入キューから前記送出キューへ前記コマンドを格納するコマンドバッチ生成手段と、

を含むことを特徴とする外部記憶装置。

【請求項2】 請求項1記載の外部記憶装置において、前記外部記憶制御部と前記記憶ディスク部はバスにより接続され、

前記外部記憶制御部は、前記記憶ディスク部に前記コマンドを試行的に発行し、そのコマンドに対し前記記憶ディスク部が前記バスの使用権を獲得するまでの応答時間を計測する応答計測手段を有し、

前記コマンドキューイング手段は、前記コマンドの発行を前記バスのフリー状態の検出から前記応答時間に応じた時間、遅延させる遅延手段を有すること、

を特徴とする外部記憶装置。

【請求項3】 請求項1記載の外部記憶装置において、前記外部記憶制御部は、

前記記憶ディスク部に対する読み出しコマンドの処理待ち数を計数するコマンド計数手段と、

複数の読み出しコマンドに対する前記記憶ディスク部からの読み出しデータを保持するバッファを有し、

前記処理待ち数に基づいて前記バッファを制御し、複数の読み出しコマンドの前記読み出しデータを前記バッファへ蓄積させ、その蓄積された読み出しデータをコンピュータへ一括転送させるバッファ制御手段と、

を含むことを特徴とする外部記憶装置。

【請求項4】 請求項1記載の外部記憶装置において、前記外部記憶制御部は、

前記各送出キューから送出され処理待ち状態にあるコマンドについての前記コマンド処理時間予測値の最大値をタイムアウト値として前記各記憶ディスク部ごとに保持するタイムアウト値保持手段と、

直前のコマンドの処理終了からの経過時間を、前記各記憶ディスク部ごとに計測する経過時間計測手段と、

前記経過時間がその対応する前記記憶ディスク部の前記タイムアウト値を越えたときエラーとして判定するエラー判定手段と、

を含むことを特徴とする外部記憶装置。

【請求項5】 請求項1記載の外部記憶装置において、前記コマンドキューイング手段は、前記コマンドによりアクセスされる前記記憶ディスク部のアドレスに基づいて、アクセス時間が最小と予測されるコマンドを選択するアクセス最適化手段を有することを特徴とする外部記憶装置。

【請求項6】 請求項5記載の外部記憶装置において、前記外部記憶制御部が複数個設けられて前記記憶ディスク部を共有し、

この各外部記憶制御部は、アクセスした前記記憶ディスク部のアドレスに基づくアドレス情報を他の外部記憶制御部に通知するアドレス情報通知手段を有し、

前記アクセス最適化手段は、前記アドレス情報を参照して前記送出キューから前記コマンドを選択すること、

を特徴とする外部記憶装置。

【請求項7】 請求項1記載の外部記憶装置において、前記記憶ディスク部は、2つのディスク装置を含み、これらディスク装置間においてミラーリングが行われるレベル1のRAIDアレイ・ディスクであり、

前記外部記憶制御部は、前記送出キューの中の前記2つのディスク装置に対する読み出しコマンドを、それらの前記コマンド処理時間予測値に基づいて前記2つのディスク装置に振り分けて発行し、これら両ディスク装置の読み出し処理時間の均等化を図るRAID1負荷分散手段を有すること、

を特徴とする外部記憶装置。

【請求項8】 請求項1記載の外部記憶装置において、前記記憶ディスク部は、複数のディスク装置を含み、これら各ディスク装置に格納されるデータに基づいて生成されるパリティをこれら全ディスク装置に分割して保持するレベル5のRAIDディスク・アレイであり、

前記外部記憶制御部は、書き込みコマンドの処理における新しいパリティの生成処理に用いるデータとして、更新されるデータと現パリティとを読み出すか、これら以外の現パリティの生成に用いたデータを読み出すかを、読み出し処理時間が前記ディスク装置間にて均等化されるように前記コマンド処理時間予測値に基づいて選択するRAID5負荷分散手段を有すること、

を特徴とする外部記憶装置。

【請求項9】 請求項1記載の外部記憶装置において、

前記記憶ディスク部は、複数のディスク装置を含み、前記各ディスク装置に格納されるデータに基づいてそのセクタ単位にパリティを生成し、このパリティを特定の前記ディスク装置に集中して保持させて高転送速度を実現するレベル 3 の RAID ディスク・アレイであり、前記外部記憶制御部は、分割された複数の前記セクタ範囲ごとに、パリティを保持する前記ディスク装置が異なるようにパリティ格納位置を生成するパリティ位置割付手段を有すること、

を特徴とする外部記憶装置。

【請求項 10】 請求項 7 から請求項 9 までのいずれかに記載の外部記憶装置において、

前記外部記憶制御部は、起動時に前記ディスク装置のシーク時間及び回転時間を測定し、この測定値に基づいて前記コマンド処理時間予測値を設定する予測処理時間設定手段を有することを特徴とする外部記憶装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、計算機システムに用いられる複数の記憶ディスク部を有した外部記憶装置に関し、特にそのレスポンスとスループットの向上による可用性の向上に関する。

【0002】

【従来の技術】ディスク装置における読み出し／書き込み処理には機械的動作が伴うため、その処理速度は、計算機の CPU 処理速度に比べると段違いに遅い。しかし、ディスク装置は、大量のデータを安価に記憶できる点で優れている。このため、現在の計算機システムにおいては、上記処理速度の格差を補う様々な工夫をした上で、ディスク装置は大規模容量を有する外部記憶装置に採用されている。

【0003】図 15 は、例えば特開平 6 - 19625 に示される従来の外部記憶装置の模式的な構成図である。この外部記憶装置 2 は、2 つの記憶ディスク部 4、6 を含み、2 つのディスク制御部 8、10 がディスク接続部 12 を介して記憶ディスク部 4、6 に接続されている。ディスク制御部 8、10 にはそれぞれホスト計算機 14、16、18 に接続され、各ホストコンピュータは一方のディスク制御部が使用中でももう一方のディスク制御部にアクセスし外部記憶装置 2 を利用することができる。各ディスク制御部はホスト計算機 14、16、18 が発行する外部記憶装置 2 に対するコマンドに基づいて、各記憶ディスク部を制御する。

【0004】ホスト計算機 14、16、18 が発行するコマンドはディスク接続部 12 内の共有メモリ 20 に設けられる待ち行列（キュー）に格納される。キュー内のコマンドはバスを介して順次、記憶ディスク部に対し送られる。記憶ディスク部がコマンドのキューイング機能を有する場合、すなわち記憶ディスク部自身が受け付けたコマンドのキューを備えている場合には、共有メモリ

内の前記キューからのコマンドの送出は、それ以前に送出されたコマンドの処理終了を待たずに行うことができる。

【0005】上記のようにディスク制御部を複数にすることにより、ディスク制御部の獲得におけるホストコンピュータ間の競合が緩和され、ホストコンピュータが外部記憶装置 2 に対してアクセスするまでの時間が短縮される。また、記憶ディスク部を複数にすることにより、1 つの記憶ディスク部がヘッドのシークによるポジショニング動作やディスクの回転動作を行う間に、ディスク制御部 8、10 と他の記憶ディスク部との間でコマンドの発行やデータ転送を行うことができ、コンピュータが実行するプログラム全体でのレスポンス時間の低減やスループットの向上が図られる。また、メモリ 20 をディスクキャッシュとして機能させてもレスポンス時間を低減することができる。

【0006】さて、外部記憶装置の処理性能を向上させるため前記キューを用いる方法として、特開平 6 - 259198 に、アクセスされる記憶ディスク部の領域を連続とすることができる場合には、それに応じてキュー中のコマンドを並べ換え、これにより上記ポジショニング動作やディスクの回転動作を効率化する方法が示されている。

【0007】

【発明が解決しようとする課題】従来は、キュー内のコマンドをどのように処理すれば外部記憶装置の処理性能が向上するかについては十分な配慮がなされているとは言えず、外部記憶装置の処理はレスポンスやスループットに関し必ずしも効率的には行われていないという問題点があった。例えば、上述の構成のように複数の記憶ディスク部がバスを介してディスク制御部に接続されている場合ではバスアービトレーション、すなわちバスの使用権の獲得等を考慮しなければ、最適な処理を行うことはできない。ここで、複数の記憶ディスク部を有する外部記憶装置における最適な処理とは、装置全体のレスポンスやスループットが優れているだけでなく、記憶ディスク部間におけるレスポンスやスループットの不均衡がないことである。

【0008】本発明は複数の記憶ディスク部を有した上記最適な処理を実現する外部記憶装置を提供することを目的とする。

【0009】

【課題を解決するための手段】本発明に係る外部記憶装置においては、外部記憶制御部は、コマンドの処理に要すると予測される時間であるコマンド処理時間予測値を生成する予測処理時間生成手段と、所定のタイミングで受入キュー内のコマンドをその前記コマンド処理時間予測値の和が所定の処理タイムスライスに応じた値となる個数だけ格納される前記コマンドの待ち行列である送出キューと、前記予測処理時間生成手段に基づき各記憶デ

ディスク部別の処理時間を予測する処理時間予測手段と、この予測された処理時間が最大の前記記憶ディスク部に対する前記コマンドを前記送出キューから取り出して対応する前記記憶ディスク部にキューイングするコマンドキューイング手段と、前記各送出キューが空になると前記受入キューから前記送出キューへ前記コマンドを格納するコマンドバッチ生成手段と、を含むことを特徴とする。

【0010】本発明によれば、送出キューが空になると、受入キューに蓄えられたコンピュータからのコマンドが一群ずつ取り出され、送出キューに振り分けられる。このとき、この一群をなすコマンドの個数は、そのコマンド処理時間予測値の和が所定の処理タイムスライスに応じた値となるように定められる。送出キューから各記憶ディスク部へのコマンドの発行は、キュー内のコマンドの前記コマンド処理時間予測値の和が最大である送出キューを選択し、そのキュー内のコマンドを所定個数（例えば1個）、取り出して送出するという動作を繰り返す。1回の送出動作はある程度の時間を要する。これは例えば、外部記憶制御部と記憶ディスク部とがバスで接続されている場合には、1回の送出動作に、外部記憶制御部によるバスアービトレーションの時間やバスの専有時間が必要であるためである。各記憶ディスク部はキューイング可能であり、複数のコマンドを受け付けることができる。よって基本的にはある記憶ディスク部に対する送出動作は、その記憶ディスク部で前に送出されたコマンドの処理を行っている最中でも可能である。前記コマンド処理時間予測値は、例えば記憶ディスク部の最大シーク時間や最大回転待ち時間などに基づいて予測処理時間生成手段により生成される。この予測処理時間生成手段は、コマンド処理時間予測値を格納したテーブルでもよい。さて、ヘッドのシークやディスクの回転といった機械的動作など各記憶ディスク部における処理時間の多くは記憶ディスク部間で独立であり、その部分は並列処理することができる。ここで上記のように処理時間長の長い記憶ディスク部に対するコマンドの処理を優先して開始することにより、この並列処理の割合が向上し、上記送出キューに振り分けられた一群のコマンドの各処理時間が短縮される。また、ある一群のコマンドの各記憶ディスク部への送出が終了してから、次の一群のコマンドの送出が開始される。そのためどのコマンドの処理も基本的に処理タイムスライス以下に終了し、最大レスポンス時間及び最低スループットが保証される。

【0011】本発明に係る外部記憶装置においては、前記外部記憶制御部と前記記憶ディスク部はバスにより接続され、前記外部記憶制御部は前記記憶ディスク部に前記コマンドを試行的に発行し、そのコマンドに対し前記記憶ディスク部が前記バスの使用権を獲得するまでの応答時間を計測する応答計測手段を有し、前記コマンドキューイング手段は前記コマンドの発行を前記バスのフリ

ー状態の検出から前記応答時間に応じた時間、遅延させる遅延手段を有すること、を特徴とする。一般に、外部記憶制御部はバス・フリー検出からアービトレーションの実行までを記憶ディスク部より短い時間で行うことができる。本発明によれば、例えば記憶ディスク部に格納されたデータに影響を及ぼさないコマンドを試行的に発行し、その記憶ディスク部からのレスポンスが外部記憶制御部に戻るまでの時間に基づいて応答時間を定める。望ましくはこの応答時間として、コマンドの試行的な発行を複数回行ってそれらの平均を採用するのがよい。この応答時間だけ外部記憶制御部から記憶ディスク部へのコマンドの発行を遅延させることにより、記憶ディスク部のバス獲得における優先度が増し、記憶ディスク部からのレスポンス時間が短縮される。

【0012】本発明に係る外部記憶装置においては、前記外部記憶制御部は、前記記憶ディスク部に対する読み出しコマンドの処理待ち数を計数するコマンド計数手段と、複数の読み出しコマンドに対する前記記憶ディスク部からの読み出しデータを保持するバッファを有し、前記処理待ち数に基づいて前記バッファを制御し複数の読み出しコマンドの前記読み出しデータを前記バッファへ蓄積させ、その蓄積された読み出しデータをコンピュータへ一括転送させるバッファ制御手段と、を含むことを特徴とする。

【0013】本発明によれば、コマンド計数手段が、受入キューから取り出された一群のコマンド中の読み出しコマンドのうち記憶ディスク部からレスポンスが帰ってきていないコマンド数である前記処理待ち数をカウントする。バッファ制御手段はこの処理待ち数に基づき記憶ディスク部から読み出されるデータがあるか否かを判断し、基本的にまだ読み出されるデータがある場合にはバッファからコンピュータへの転送を見合わせて次の読み出しデータの到着を待ち、一方、もう読み出されるデータがない場合やバッファの残り空き容量が少ない場合には、バッファからコンピュータへの転送を行う。このように複数の読み出しコマンドに対応するデータを一括してコンピュータへ転送することにより、その転送の際の外部記憶装置からコンピュータへのチャネルの接続処理といった時間的オーバーヘッドが低減され、レスポンスやスループットの向上が図られる。

【0014】本発明に係る外部記憶装置においては、前記外部記憶制御部は、前記各送出キューから送出され処理待ち状態にあるコマンドについての前記コマンド処理時間予測値の最大値をタイムアウト値として前記各記憶ディスク部ごとに保持するタイムアウト値保持手段と、直前のコマンドの処理終了からの経過時間を、前記各記憶ディスク部ごとに計測する経過時間計測手段と、前記経過時間がその対応する前記記憶ディスク部の前記タイムアウト値を越えたときエラーとして判定するエラー判定手段と、を含むことを特徴とする。

【0015】本発明によれば、前記経過時間はコマンドの処理が終了すればリセットされるので、経過時間はコマンドの処理が正常に行われているならば対応するタイムアウト値よりも小さい値である。一方、前記タイムアウト値は前記各送出キューから送出され記憶ディスク部にキューイングされたコマンドのうち未処理のコマンドについての前記コマンド処理時間予測値の最大値であるので、経過時間がタイムアウト値より大きくなるということは、それら未処理コマンドのうちいずれのコマンドも終了しないという異常を示している。エラー判定手段は、この経過処理時間がタイムアウト値を超えたタイミングにおいてエラーを検出する。よって、エラー判定手段は、前記各送出キューの予測された処理時間の終了より早くエラーを検出することができる。

【0016】本発明に係る外部記憶装置においては、前記コマンドキューイング手段は、前記コマンドによりアクセスされる前記記憶ディスク部のアドレスに基づいて、アクセス時間が最小と予測されるコマンドを選択するアクセス最適化手段を有することを特徴とする。

【0017】本発明によれば、アクセス最適化手段は、前記送出キュー内の前記コマンドにより指定される前記アドレスが例えば昇順、降順となるようにコマンドを選択し、これらコマンドの実行における記憶ディスク部のヘッドのシーク時間やディスクの回転時間といったアクセス時間を最小とする。これによりレスポンスやスループットの向上が図られる。

【0018】本発明に係る外部記憶装置においては、前記外部記憶制御部が複数個設けられて前記記憶ディスク部を共有し、この各外部記憶制御部はアクセスした前記記憶ディスク部のアドレスに基づくアドレス情報を他の外部記憶制御部に通知するアドレス情報通知手段を有し、前記アクセス最適化手段は前記アドレス情報を参照して前記送出キューから前記コマンドを選択すること、を特徴とする。

【0019】本発明によれば、アドレス情報通知手段が、一方の外部記憶制御部において最後にアクセスされたアドレスに関する情報を、次の処理を行う外部記憶制御部に通知する。次の処理を行う外部記憶制御部はこのアドレス情報に基づいて、例えば前記アクセス時間が最小であるコマンドから処理を開始する。これによりレスポンスやスループットの向上が図られる。

【0020】本発明に係る外部記憶装置においては、前記記憶ディスク部は2つのディスク装置を含みこれらディスク装置間においてミラーリングが行われるレベル1のRAIDアレイ・ディスクであり、前記外部記憶制御部は前記送出キューの中の前記2つのディスク装置に対する読み出しコマンドをそれらの前記コマンド処理時間予測値に基づいて前記2つのディスク装置に振り分けて発行しこれら両ディスク装置の読み出し処理時間の均等化を図るRAID1負荷分散手段を有すること、を特徴

とする。レベル1のRAIDアレイ・ディスクでは、読み出し処理は一方のディスク装置に対してのみ行えばよい。本発明によれば、送出キュー中に複数の読み出しコマンドがあれば、それらをレベル1のRAIDアレイ・ディスクを構成する2つのディスク装置に振り分けて発行し、これら両ディスク装置において異なる読み出しコマンドを並列に処理させることにより、コマンド処理の効率が向上する。

【0021】本発明に係る外部記憶装置においては、前記記憶ディスク部は複数のディスク装置を含みこれら各ディスク装置に格納されるデータに基づいて生成されるパリティをこれら全ディスク装置に分割して保持するレベル5のRAIDディスク・アレイであり、前記外部記憶制御部は書き込みコマンドの処理における新しいパリティの生成処理に用いるデータとして、更新されるデータと現パリティとを読み出すかこれら以外の現パリティの生成に用いたデータを読み出すかを、読み出し処理時間が前記ディスク装置間にて均等化されるように前記コマンド処理時間予測値に基づいて選択するRAID5負荷分散手段を有すること、を特徴とする。

【0022】レベル5のRAIDディスク・アレイでは、データの書き込み処理はパリティの更新処理を伴う。例えば n 個のデータを D_k ($k=1\sim n$)、これらにより生成されるパリティを P とする。これらはレベル5のRAIDディスク・アレイを構成する各ディスク装置に分散して保持されている。ここで D_j ($1\leq j\leq n$)を D_j' で更新する書き込み処理を考える。新たな n 個のデータに対するパリティを求める第1の方法は D_j' 、 D_j 及び P から求める方法であり、第2の方法は D_j' 及び D_k ($k\neq j$)から求める方法である。この第1の方法においては D_j 及び P をディスク装置から読み出さなければならない。一方、第2の方法においては D_k ($k\neq j$)を読み出さなければならない。この第1の方法と第2の方法とでは、読み出し処理が行われるディスク装置が異なる。本発明によれば、RAID5負荷分散手段は送出キュー中に複数の書き込みコマンドがあれば、それらに伴う上記パリティ更新における読み出し処理が、各ディスク装置に分散し処理時間が均等化するように、上記第1の方法、第2の方法を選択する。すなわち書き込みコマンドに伴う読み出し処理が各ディスク装置間において並列に処理されることにより、コマンド処理の効率が向上する。

【0023】本発明に係る外部記憶装置においては、前記記憶ディスク部は複数のディスク装置を含み前記各ディスク装置に格納されるデータに基づいてそのセクタ単位にパリティを生成しこのパリティを特定の前記ディスク装置に集中して保持させて高転送速度を実現するレベル3のRAIDディスク・アレイであり、前記外部記憶制御部は分割された複数の前記セクタ範囲ごとに、パリティを保持する前記ディスク装置が異なるようにパリテ

ィ格納位置を生成するパリティ位置割付手段を有すること、を特徴とする。

【0024】本発明によれば、例えば m 個のディスク装置から構成されるレベル3のRAIDディスク・アレイにおいて、データは m 個のディスク装置に格納される。これにより、従来の1つのディスク装置がパリティ専用であって残りの $(m-1)$ 個がデータを格納するディスク・アレイよりも、データの分散度合いが高まり、コマンド処理が各ディスク装置間において並列に処理される可能性が高くなり、コマンド処理の効率が向上する。

【0025】本発明に係る外部記憶装置においては、前記外部記憶制御部は起動時に前記ディスク装置のシーク時間及び回転時間を測定しこの測定値に基づいて前記コマンド処理時間予測値を設定する予測処理時間設定手段を有することを特徴とする。本発明によれば、シーク時間及び回転時間の実測値を利用することにより、コマンド処理時間予測値の精度が高まり、外部記憶制御部によるコマンド処理の効率向上のための制御精度が上がる。

【0026】

【発明の実施の形態】以下、本発明の好適な実施形態を

図面を参照して説明する。
【0027】〔実施形態1〕図1は、本発明が適用される共有外部記憶装置の模式的な構成図である。共有外部記憶装置30は、ホスト計算機32、34、36で共有される外部記憶装置である。共有外部記憶装置30は、データを格納する3つのディスク装置38（ディスク装置A、B、C）を有し、これらがそれぞれ記憶ディスク部を構成する。また共有外部記憶装置30は、外部記憶制御部40を有し、これが各ホスト計算機とディスク装置38との間のデータの入出力を制御する。ディスク装置38と外部記憶制御部40とはSCSIバス42により互いに接続されている。なお、各ホスト計算機と外部記憶制御部40との間もそれぞれSCSIバス44により接続されている。

【0028】図2は共有外部記憶装置30の内部処理を説明するためのブロック図である。外部記憶制御部40内の処理を中心に説明する。受入キュー50が、ホスト計算機32、34、36から発行された読み出しコマンド又は書き込みコマンドを一括して受付て蓄積する。本外部記憶制御部40は予測処理時間生成手段として、コマンド処理時間予測値を格納した予測値テーブル52を有している。このコマンド処理時間予測値は、ディスク装置38の最大シーク時間、最大回転待ち時間、ディスク装置38内部のデータ転送レート、SCSIバス42のデータ転送レート等の性能値から予測したコマンドの実行時間であり、例えば読み出し／書き込みの別や転送されるデータ量などによって異なる。

【0029】コマンドバッチ生成部54は、予測値テーブル52に基づいて受入キュー50内のコマンドの実行時間を予測し、コマンド処理時間予測値の和が所定のバ

ッチ時間 t を超えない範囲でできる限り多くのコマンド56を受入キュー50の先頭から読み込む。コマンドバッチ生成部54は、読み込んだ一群（バッチ）のコマンドの処理が終了するまで、次の読み出し動作を行わない。すなわち、コマンドバッチ生成部54は、受入キュー50をバッチ処理する。コマンド56はコマンドバッチ生成部54から送出キュー58に格納される。

【0030】コマンドキューイング部60は後述するやり方で送出キュー58から1つずつコマンドをSCSIバス42に送出する。各ディスク装置38はディスク側キュー62を有しコマンド・キューイングをサポートしている。よってコマンドキューイング部60はSCSIバス42がフリーであれば、先に送出したコマンドのレスポンスがディスク装置38から返っていなくても、同一のディスク装置38に対して多重にコマンドを発行することができる。各ディスク装置38は、自身に多重に発行されたコマンドをディスク側キュー62に蓄積し、これから順次読み出して実行する。実行された結果、例えば読み出しデータはディスク装置38からSCSIバス42を介して、外部記憶制御部40に送られるが、この戻りの経路は図示していない。

【0031】図3は、コマンドキューイング部60の処理を説明するための送出キューの模式図である。コマンドバッチ生成部54は受入キュー50から上述したようにバッチ時間 t を基準として1群のコマンドを読み出す。図3に示す例は、10個のコマンドが読み出され、そのうち4個がディスク装置A、1個がディスク装置B、5個がディスク装置Cを対象とするものであり、これらは送出キュー58に格納される。キュー70、72、74はそれぞれディスク装置A、B、Cに対応する処理待ちコマンドの列を表している。これらキューに含まれる1～10の番号を付したブロックは各コマンドを表し、これら各ブロックの高さは各コマンドのコマンド処理時間予測値を表す。よって各キューの縦方向の長さは各記憶ディスク部における処理時間の予測値である。

【0032】コマンドキューイング部60は、送出キュー58から処理時間の予測値が最長の記憶ディスク部に対応するコマンドを1つ選択して読み出し、SCSIバスを介して、対応するディスク装置38に送出する。例えば、図3においてディスク装置Cに対応するキュー74が最長であるので、その中のコマンド（番号“1”で表す）を送出する。この状態で、まだ同じキュー74が最長なので、このキュー74からコマンド“2”を送出する。すると今度は、ディスク装置Aに対応するキュー70が最長となるので、そのキュー70からコマンド“3”を送出する。同様に続いてコマンド“4”～“10”が順次送出される。なお、これら各キュー間の比較はコマンド処理時間予測値によるものであり、コマンド数には依存しない。例えば、コマンド“5”まで送出した時点で、キュー70、74にはコマンドがそれぞれ2

個残っており、キュー72には1個しか残っていない。しかし処理時間はキュー72の方が他の2つのキューよりも長いので、他のキューより優先してキュー72のコマンド“6”が送出される。

【0033】図4、図5は上述したコマンドキューイングの効果を説明するための図である。図4は、ディスク装置38を2つとした場合における送出キュー内のコマンド列の1例を示す模式図である。図5は、図4に示した例における処理のタイミングチャートである。図4に示す例では、ディスク装置Aに対応するキュー80は3個のコマンドa、b、cを有し、ディスク装置Bに対応するキュー82は1個のコマンドを有する。コマンドa、b、c、dはそれぞれコマンド処理時間予測値として 2τ 、 4τ 、 6τ 、 7τ を有する（ τ は任意の時間単位）。なお、コマンドをディスク装置38に送る際、及びディスク装置38からコマンドの処理結果を返す際の、SCSIバスの使用権の獲得からその開放までの時間をそれぞれセンド時間、リターン時間と称することとする。センド時間、リターン時間はこの例では τ である。図5（a）は本実施形態のコマンドキューイング部の処理のタイミングチャートを示す。一方、図5（b）は、本実施形態のコマンドキューイング部とは異なる処理のタイミングチャートを示す。コマンドの対応するディスク装置への送出の順序は、図5（a）に示す本実施形態の場合においてはコマンドc、d、b、a（それぞれセンド時間90、92、94、96に対応）の順であり、一方、図5（b）に示す例ではコマンドd、c、b、a（それぞれセンド時間100、102、104、106に対応）の順である。送出されたコマンドはディスク装置において実行された後、リターン時間110、112、114、116、120、122、124、126において、外部記憶制御部40に回答が返される。

【0034】本実施形態では、2つのキューのうち短い処理時間を有するキュー82の処理期間（コマンドdの処理期間）は処理時間の長い方のキュー80の処理期間（コマンドc、b、aの処理期間）に包含される。このように、上記コマンドキューイング部の処理によれば、短い処理時間を有する記憶ディスク部の処理期間は、それより長い処理時間を有する記憶ディスク部の処理期間の背後に隠されるように、処理がスケジューリングされる。これに対し、本実施形態を異なる図5（b）に示す方法では、短い処理時間を有する記憶ディスク部の処理期間は、それより長い処理時間を有する記憶ディスク部の処理期間からはみ出す。図5（a）、（b）に示す例では処理開始から終了までの時間は、それぞれ 14τ 、 15τ である。このように本発明の実施形態においては、全記憶ディスク部に対するコマンド処理終了までの時間が短縮され、レスポンス及びスループットが向上する。

【0035】本実施形態の他の特徴を次に説明する。コ

マンドの処理に際して、上述のようにSCSIバスの使用に関するセンド時間、リターン時間が必要になる。これらの時間中には、外部記憶制御部40やディスク装置38がバス42の使用権を獲得するアービトレーションという動作を実行する必要がある。しかしディスク装置38がバス・フリー状態を検出してアービトレーションを開始するまでの時間は、外部記憶制御部40のその時間よりも遅い。さらに、ディスク装置38のSCSI IDの優先順位は、外部記憶制御部40のそれよりも低い値に設定されている。そのため、外部記憶制御部40が連続的にコマンドをディスク装置38のいずれかに送出しようとする場合、ディスク装置38側は、バス42の使用権を獲得することが困難となり、ディスク装置38からのレスポンス時間が長くなる可能性がある。

【0036】図6、図7は、本装置の外部記憶制御部40における上記問題を解決するための上記方法を説明するフローチャートである。図6は、本装置の電源オン時における自己診断での動作を説明するフローチャートである。外部記憶制御部40は、本装置の電源オン時の自己診断の一環として、バス・フリー状態を検出し試行的にコマンドを発行して、ディスク装置38がこのコマンドに対しレスポンスを返すためにバス42の使用権を獲得するまでの時間を測定する。ここで発行されるコマンドは、ディスクコネクトを要求するがディスク装置38内のデータに影響を及ぼさないコマンドであり、例えばINQUIRY、READ、CAPACITY等である。外部記憶制御部40はこれらのコマンドを複数回発行して、それらに対してディスク装置38からのレスポンスを得るまでの応答時間の平均値を計測する。この応答時間の平均値は、コマンドキューイング部60に与えられる。図7は実際のコマンド発行処理におけるコマンドキューイング部60の動作を説明するフローチャートである。コマンドキューイング部60はコマンドに対するアービトレーションの開始をバス・フリー状態の検出から前記応答時間の平均値に応じた時間、例えば、平均値に安全係数を乗じた時間だけ遅延させる。ちなみに、安全係数はデフォルト値として例えば2を与える。なお、アービトレーションの実行を遅延させている間にバス・ビジー状態となった場合には、バス・フリー状態となるのを待って再度処理を試みる処理を行う。

【0037】上記処理により、ディスク装置38のバス獲得における優先度が増し、外部記憶制御部40がディスク装置38からのレスポンスを迅速に受け取ることができる。また、遅延時間はディスク装置38ごとに変えることができ、バス・フリー状態の検出からアービトレーションの開始までの時間がディスク装置38ごとに異なる場合でも対応可能である。

【0038】また本装置では、複数の読み出しコマンドに対するホスト計算機32へのデータ転送処理を一括して行う。これによるSCSIバス44を介したホスト計

算機 32 とのデータ転送の処理時間は、その処理をコマンドごとに行う場合より短くなり、ホスト計算機 32 に対するレスポンスが向上する。この一括転送処理を以下に説明する。外部記憶制御部 40 は、1 回のコマンドのバッチ中に含まれる読み出しコマンドのうち処理待ち状態にあるコマンド数を計数する機能と、ディスク装置 38 からの読み出しデータを蓄積することができるバッファとを有する。外部記憶制御部 40 は、バッファにできるだけ多くの読み出しコマンドに対する読み出しデータを蓄積し、バッファが一杯になると、バッファからホスト計算機 32 への転送処理を開始する。これにより、ホスト計算機 32 とのバスの接続処理といったオーバーヘッドが減るのでホスト計算機 32 に対するレスポンスが向上する。前記処理待ちの読み出しコマンドの数が 0 になれば、バッファが一杯にならなくても、外部記憶制御部 40 はバッファ内のデータをホスト計算機 32 に転送する。また、外部記憶制御部 40 は、コマンド処理時間予測値に基づいてコマンド処理の予測スケジュールを作成することができるので、このスケジュールに基づいて読み出しコマンドに対するディスク装置からの応答の間隔を予測することができる。外部記憶制御部 40 は、その間隔とホスト計算機に対する転送処理のオーバーヘッド時間との比較に基づいて、バッファに次の読み出しコマンドのデータが到着するのを待つことの得失を判断して、バッファからの転送処理を制御してもよい。

【0039】本装置は、なんらかのエラーにより発行されたコマンドに対するレスポンスがない場合のタイムアウト処理を以下のように行う。ディスク装置 38 は、そのディスク側キュー 62 内のコマンドを例えばアクセスされるアドレスに基づいて最適な順序で実行する。外部記憶制御部 40 は、ディスク装置 38 側にキューイング済みであり処理待ちであるコマンドのコマンド処理時間予測値の最大値をタイムアウト値として保持するとともに、直前のコマンドの処理終了からの経過時間を計測する。

【0040】図 8 はタイムアウト処理を説明する模式図である。図 8 (a) はある 1 つのディスク装置 38 に 4 つのコマンド A、B、C、D を発行した場合の、コマンド処理時間予測値によるディスク装置の占有時間のタイムチャートである。コマンド処理時間予測値の大小関係は、 $A < C < D < B$ であるとする。外部記憶制御部 40 は、ディスク装置において最初に処理されるコマンド A の処理が完了し、そのレスポンスが返された時点から経過時間を計測し始める。またこれと同時にこのディスク装置に対するタイムアウト値は、処理待ちコマンド B、C、D のうち最長のコマンド処理時間予測値を有するコマンド B のその予測値を設定される (図 8 (b))。次にコマンド B の処理が開始される。ディスク装置はコマンド B の処理をタイムアウト値以内に完了し、外部記憶制御部 40 にレスポンスを返す (図 8 (c))。外部記

憶制御部 40 はこの時点においてはディスク装置 38 が正常動作を行っている判断し、経過時間をリセットし、またタイムアウト値をコマンド D のコマンド処理時間予測値に変更する (図 8 (d))。

【0041】図 8 (a) に示すコマンド実行のスケジュールはディスク装置により決定され、外部記憶制御部 40 はその実行順序を知らない。よって、外部記憶制御部 40 はコマンド B のレスポンスがあった時点で、次にコマンド C、D のいずれが実行されるかを知らない。しかし、タイムアウト値は、まだ実行されていないコマンドのうち最長のコマンド処理時間予測値である。よって、外部記憶制御部 40 は新たに設定されたタイムアウト値以内になんらのレスポンスもディスク装置から得られなければ、コマンド C、D のいずれも実行されずエラーと判断することができる (図 8 (e))。

【0042】このエラーが検出されるまでの時間は、コマンド A、B の処理に要した実時間とコマンド D のコマンド処理時間予測値との和である。もしコマンドのバッチ単位でタイムアウトを監視したならば、エラーはディスク装置の占有時間 (コマンド A、B、C、D のコマンド処理時間予測値の和) が経過しないと検出されない。よって上述のようにコマンド単位でタイムアウトを監視することにより、タイムアウトの検出が早まる。これによりエラーに対する対応がそれだけ早く行われ、外部記憶装置のレスポンスの向上を図ることができる。

【0043】さらに、外部記憶制御部 40 は、バッチ (コマンドのグループ) の処理が最適に行われるように、コマンド発行のスケジューリングを行う。コマンド・キューイング機能を有するディスク装置 38 は、それ自身がスケジューリング機能を有している。このディスク装置 38 におけるスケジューリングは、ディスク装置の回転待ち時間、ヘッドのシーク時間、不良セクタの置き換えなどを考慮して行われる。これらの点は外部記憶制御部 40 では把握困難であるので、これらの点に基づくスケジューリングはディスク装置側で行うのがよい。

【0044】しかし、ディスク装置側は、そのディスク側キュー 62 に蓄積されたコマンドの範囲内でスケジューリングを行い、外部記憶制御部 40 から未発行でディスク側キュー 62 に無いコマンドをスケジューリングに際して考慮することができないので、バッチ単位で見た場合、必ずしも最適な順序で処理が行われていない可能性がある。例えば、ディスク側キュー 62 が空であるときに受け取ったコマンドは、無条件で実行開始されるため、シーク時間や回転待ち時間が多くかかるコマンドから処理が行われる可能性がある。

【0045】そこで、外部記憶制御部 40 は送出キュー 58 内のコマンドを各ディスク装置ごとにそのアクセスする論理アドレスに基づき並べ換えて、順に各ディスク装置に発行する。このとき、コマンド処理時間の予測値が最長であるディスク装置に対するコマンドを取り出す

ことは上述した通りである。また、バッチが連続する場合には、後のバッチの処理は、アドレスに基づいて前のバッチの最後に実行したコマンドに対して実行時間が短いと予想されるコマンドから始める。具体的には、コマンドの発行順序をアクセスする論理アドレスの大きい順もしくは小さい順に発行する。また例えば、前のバッチが論理アドレスの大きい順にコマンドを発行した場合

は、それに続くバッチは論理アドレスの小さい順にコマンドを発行する。このように外部記憶制御部40が論理アドレスに基づいてアクセス時間が最適となるようにコマンドを発行することにより、ディスク装置38がディスク側キュー62にのみ基づいてスケジューリングを行う場合の上記問題が解決され、外部記憶装置のレスポンスが向上する。

【0046】【実施形態2】図9は、本発明が適用される他の共有外部記憶装置の模式的な構成図である。図中、特に断らない限り、図1の符号に200を加えた構成要素は、図1と同一の機能を有するものとする。本実施形態と実施形態1との主たる相違点は、本実施形態の外部記憶装置230は2つの外部記憶制御部250、252を有し、これらがディスク装置238を共有して制御する点にある。外部記憶制御部250、252及び2つのディスク装置238（ディスク装置A、B）はSCSIバス254にて接続される。両外部記憶制御部間には、それらの間での通信を可能にするバス256が設けられている。両外部記憶制御部は、それぞれ実施形態1の外部記憶制御部40と基本的には同一の動作を行う。外部記憶制御部250、252が、外部記憶制御部40と異なるのは、一方の外部記憶制御部からバッチを処理した後、続いて他のバッチをもう一方の外部記憶制御部から処理する場合である。これは、外部記憶制御部を2つ有する本装置に特有である。

【0047】本装置では、一方の外部記憶制御部が処理するバッチにおける最後のコマンドのアクセス先の論理アドレスが、ディスク装置のアドレスの最小値、最大値のいずれに近いかを、バス256を介して他方の外部記憶制御部に通知する。すなわち、バス256はアドレス情報通知手段として機能する。例えば、外部記憶制御部250は各ディスク装置238ごとに、その最後のコマンドの論理アドレスが0に近い場合には、バス256にL（low）レベルの制御信号を送出し、逆に論理アドレスが最大値に近い場合にはH（high）レベルの制御信号を送出する。例えば、外部記憶制御部252はバス256からディスク装置AについてLレベル、ディスク装置BについてHレベルの制御信号を受けた場合には、そのディスクAに対しては、そのアクセスする論理アドレスの小さいコマンドから順に発行し、ディスクBに対しては、そのアクセスする論理アドレスの大きいコマンドから順に発行する。

【0048】このように2つの外部記憶制御部250、

252間においても、コマンドが論理アドレスに基づいてアクセス時間が最適となるよう発行され外部記憶装置のレスポンスが向上する。

【0049】【実施形態3】図10は、レベル1のRAIDディスク・アレイを備え本発明が適用される共有外部記憶装置の模式的な構成図である。図中、特に断らない限り、図9の符号に100を加えた構成要素は、図9と同一の機能を有するものとする。本実施形態と実施形態2との主たる相違点は、本実施形態の外部記憶装置330は2つのレベル1のRAIDディスク・アレイ（RAID1ディスク・アレイ）358、360を記憶ディスク部として有する点にある。各RAID1ディスク・アレイは2つのディスク装置から構成され、これらディスク装置間でミラーリングが行われる。各RAID1ディスク・アレイを構成する2つのディスク装置は互いに異なるSCSIバス362、364に接続される。これら2つのSCSIバス362、364は、外部記憶制御部350、352にもそれぞれ接続される。両外部記憶制御部350、352は基本的には、各RAID1ディスク・アレイをそれぞれ1個の記憶ディスク部として扱い、実施形態2と同様の制御を行う。但し、本装置においては、読み出し処理における最適化を行う点が、上記実施形態と異なる。以下、この点を、RAID1ディスク・アレイ360を例にとって説明する。

【0050】RAID1ディスク・アレイ360では、書き込み処理は、同一のデータを2つのディスク装置366、368の双方に書き込まなければならない。しかし、読み出し処理については、ディスク装置366、368のいずれか一方から読み出せばよい。そこで、外部記憶制御部350、352は、コマンド処理時間予測値に基づいて、RAID1ディスク・アレイ360に対する読み出し処理をディスク装置366、368に振り分け、並列処理させる。これにより、コマンドの全体の処理時間が短縮され、装置のレスポンスが向上する。

【0051】表1は、上記処理を説明するためのRAID1ディスク・アレイ360に対するコマンド群の例を表しており、これらは外部記憶制御部350から発行されるものとする。表にはコマンドの読み出し／書き込みの種別とコマンド処理時間予測値である予測実行時間が示されている。

【0052】

【表1】

	コマンド	予測実行時間
a	READ	30ms
b	READ	120ms
c	WRITE	40ms
d	READ	40ms
e	WRITE	35ms
f	READ	60ms

コマンドc、eは書き込みコマンドであるので、ディスク装置366、368の双方に発行される。残りのコマンドa、b、d、fは読み出しコマンドであるので、ディスク装置366、368のいずれかに発行されればよい。外部記憶制御部350は、これら読み出しコマンドをそのコマンド処理時間予測値の和がなるべく均等化するように2つのグループに分ける。ここでは、コマンドa、d、fとコマンドbとに分ける。これらは各グループは120mSのコマンド処理時間予測値の和を有する。この外部記憶制御部350は、コマンドa、d、fをディスク装置366に発行し、コマンドbをディスク装置368に発行する。

【0053】上記の処理はRAIDタスクにて行われる。RAIDタスクはコマンドパッチ生成手段であるモニタにより生成されたコマンド群を受け取り、それを上記の手法により効率的にディスク装置に割り振る。

【0054】【実施形態4】図11は、レベル5のRAIDディスク・アレイを備え本発明が適用される共有外部記憶装置の模式的な構成図である。図中、特に断らない限り、図10の符号に100を加えた構成要素は、図10と同一の機能を有するものとする。本実施形態と実施形態3との主たる相違点は、本実施形態の外部記憶装置430は2つのレベル5のRAIDディスク・アレイ(RAID5ディスク・アレイ)458、460を記憶ディスク部として有する点にある。各RAID5ディスク・アレイは5つのディスク装置から構成され、これらディスク装置の格納するデータから生成されるパリティは、ディスク装置間で分割して保持される。各RAID5ディスク・アレイを構成する5つのディスク装置は互いに異なるSCSIバス462、464、466、468、470に接続される。これら5つのSCSIバスは、外部記憶制御部450、452にもそれぞれ接続される。両外部記憶制御部450、452は基本的には、各RAID5ディスク・アレイをそれぞれ1個の記憶ディスク部として扱い、実施形態2と同様の制御を行う。但し、本装置においては、パリティの書き込み処理にお

*ける最適化を行う点が、上記実施形態と異なる。以下、この点を、RAID5ディスク・アレイ460を例として説明する。

【0055】RAID5ディスク・アレイ460では、データの書き込み処理はパリティの更新処理を伴う。例えばn個のデータを D_k ($k=1\sim n$)、これらにより生成されるパリティをPとする。これらはレベル5のRAIDディスク・アレイを構成する各ディスク装置に分散して保持されている。ここで D_j ($1\leq j\leq n$)を D_j' で更新する書き込み処理を考える。新たなn個のデータに対するパリティを求める第1の方法は D_j' 、 D_j 及びPから求める方法であり、第2の方法は D_j' 及び D_k ($k\neq j$)から求める方法である。この第1の方法においては D_j 及びPをディスク装置から読み出さなければならない。一方、第2の方法においては D_k ($k\neq j$)を読み出さなければならない。この第1の方法と第2の方法とでは、読み出し処理が行われるディスク装置が異なる。そこで外部記憶制御部450、452は、RAID5ディスク・アレイ460に対する複数の書き込みコマンドについての上記パリティ更新における読み出し処理が各ディスク装置に分散し処理時間が均等化するように、上記第1の方法、第2の方法を選択する。すなわち書き込みコマンドに伴う読み出し処理が各ディスク装置間において並列に処理される。これにより、コマンドの全体の処理時間が短縮され、装置のレスポンスが向上する。

【0056】表2は、上記処理を説明するためのRAID5ディスク・アレイ460に対するコマンド群の例を表しており、これらは外部記憶制御部450から発行されるものとする。表にはコマンドの読み出し/書き込みの種別、アクセス先の論理アドレス、データ長及び予測実行時間が示されている。ここで書き込みコマンドの予測実行時間は、パリティ計算のための読み出し動作に必要な時間を含まないコマンド処理時間予測値である。

【0057】

【表2】

	コマンド	論理アドレス	データ長(ブロック数)	予想実行時間
a	READ	2543	1	31ms
b	WRITE	324	1	62ms
c	WRITE	686	1	62ms
d	READ	21	1	31ms
e	WRITE	760	1	62ms
f	WRITE	5890	1	62ms

コマンド処理時間予測値を求める際に、ディスク装置472、474、476、478、480(ディスク装置A、B、C、D、E)に関して、最大シーク時間は18mS、最大回転待ち時間は12mS、内部データ転送レートは5MB/S及びコントローラのオーバーヘッドは0.7mSであるという条件を使用し、また1ブロック

のデータ長は1KB、SCSIバス462、464、466、468、470の転送レートは10MB/Sであるとした。

【0058】図12はRAID5ディスク・アレイ460のセクタ・アドレスの割付けマップを表す模式図である。セクタ・アドレスは図12に示されるような規則性

に基づいて割り付けられる。この割付け方と表2の各コマンドのアドレスとに基づいて、各ディスク装置ごとの処理時間の予測値を求めると、ディスク装置A、B、C、D、Eそれぞれについて、93mS、93mS、0mS、62mS、62mSとなる。

【0059】さて、コマンドb、c、e、fは書き込みコマンドであるので、上述のようにバリティ計算のため、ディスク装置に対する読み出し動作を伴う。例えば、コマンドbに対する上述の第1のバリティ生成方法は、ディスク装置Bの古いデータとディスク装置Aの古いバリティとを読み出す動作を伴い、また第2のバリティ生成方法は、ディスク装置C、D、Eの対応セクタの古いデータを読み出す動作を伴う。バリティ計算のための読み出し動作を含まないディスク装置の占有時間(表2)に基づく比較では、ディスク装置A、Bの負荷が高くなっている。そこでコマンドbのバリティ生成処理は上記第2の方法により行い、ディスク装置間でその占有時間を平均化する。コマンドc、e、fについても上記第1、第2の方法のいずれかを、ディスク装置間の処理の均等化が図られるように選択する。これによりコマンド全体の処理が効率化され、外部記憶装置430のレスポンスが向上する。

【0060】上記の処理はRAIDタスクにて行われる。RAIDタスクはコマンドバッチ生成手段であるモニタにより生成されたコマンド群を受け取り、それを上記の手法により効率的にディスク装置に割り振る。

【0061】また、上では、コマンド処理時間予測値を、接続されるディスク装置の最大シーク時間、最大回転待ち時間等を基に算出しているが、これらの値に代えて、予め測定により得た評価値を用いて算出することもできる。測定値を取り入れることで、コマンド処理時間予測値の精度が向上し、外部記憶制御部の制御精度が向上する。これにより、装置のレスポンスの向上も期待できる。

【0062】具体的には、外部記憶制御部450、452は電源オン時に自己診断として、SEEKコマンドを複数回発行する。そして、その際にシーク・回転時間を測定し、それぞれの平均値を算出し、その値に安全係数を掛けた値をシーク時間、回転待ち時間の評価値とする。この評価値に基づきコマンドの実行時間を予測することにより、最大値を使用した場合より精度のよい予測が可能となり、処理の効率向上が図られる。なお安全係数は2をデフォルト値とし、また、システムごとに選択・変更可能とする。

【0063】【実施形態5】模式的な構成は図11と同一の装置であって記憶ディスク部458、460をレベル3のRAIDディスク・アレイで置換した装置にも本発明は適用される。この場合を図11を流用して、実施形態4と異なる点のみを説明する。本実施形態と実施形態4との主たる相違点は、記憶ディスク部内のデータの

持ち方にある。図13は、従来のRAID3ディスク・アレイのセクタ・アドレスの割付けマップを表す模式図である。従来においては、1つのディスク装置(図ではディスク装置E)がバリティ専用に割り当てられていた。これに対し、図14は本装置におけるRAID3ディスク・アレイのセクタ・アドレスの割付けマップを表す模式図である。本装置では、バリティを格納するディスク装置は、128セクタごとになる。すなわち、バリティは、0~127セクタ(0~511バイトに相当)ではディスク装置Eに格納され、次の128~255セクタ(512~1023バイト)ではディスク装置Aに格納され、以降128セクタごとにディスク装置B、C、D、…に格納される。

【0064】ここで、外部記憶制御装置が5ブロック分(0~2560バイト)のデータを読み出すコマンドを発行した場合を説明する。既存のRAID3ディスク・アレイでは、データが格納されている4つのディスク装置A~Dのそれぞれに640回のリード・アクセスを行う必要があった。ところが、本装置では、セクタ単位でバリティが割り振られているため、5つのディスク装置A~Eのそれぞれに512回のリード・アクセスを行うこととなる。これにより、アクセスがディスク装置間にて均等化され、20%性能が向上する。この方式は、1度に大量のデータ転送を必要とする画像データ等のマルチメディアアプリケーション用データの入出力において特に効果を発揮する。

【0065】

【発明の効果】本発明の外部記憶装置によれば、複数の記憶ディスク部を有する外部記憶装置において、装置全体のレスポンスやスループットが優れているとともに、記憶ディスク部間におけるレスポンスやスループットの不均衡がないという効果がある。コマンドの処理をバッチごとに進めることにより最大レスポンス時間、最低スループットが保証された外部記憶装置が得られるという効果がある。

【0066】また、本発明の外部記憶装置によれば、バス獲得における優先度に関し、外部記憶制御部と記憶ディスク部との間の均衡がとれ、記憶ディスク部からのレスポンス時間が短縮され、装置のレスポンスが向上するという効果がある。

【0067】本発明の外部記憶装置によれば、複数の読み出しコマンドに対応するデータが一括してコンピュータへ転送されるので、その転送の際の外部記憶装置からコンピュータへのチャンネルの接続処理といった時間的オーバーヘッドが低減され、レスポンスやスループットが向上するという効果が得られる。

【0068】本発明の外部記憶装置によれば、コマンド単位でタイムアウトを監視することにより、タイムアウトの検出が早まる。これによりエラーに対する対応がそれだけ早く行われ、外部記憶装置のレスポンスの向上を

図ることができる。

【0069】本発明の外部記憶装置によれば、送出キュー内の前記コマンドにより指定される前記アドレスが例えば昇順、降順となるようにコマンドが選択され、これらコマンドの実行における記憶ディスク部のヘッドのシーク時間やディスクの回転時間といったアクセス時間が最小となる。これによりレスポンスやスループットが向上するという効果が得られる。

【0070】本発明の外部記憶装置によれば、複数の外部記憶制御部を有する場合、一方の外部記憶制御部に
10 いて最後にアクセスされたアドレスに関する情報を、次の処理を行う外部記憶制御部に通知する。次の処理を行う外部記憶制御部はこのアドレス情報に基づいて、例えば前記アクセス時間が最小であるコマンドから処理を開始するので、レスポンスやスループットが向上するという効果が得られる。

【0071】本発明の外部記憶装置によれば、送出キュー中に複数の読み出しコマンドがあれば、それらをRAID1アレ
10 ID1ディスクを構成する2つのディスク装置に振り分けて発行し、これら両ディスク装置において異なる読み出しコマンドを並列に処理させることにより、コマンド処理の効率が向上する。

【0072】本発明の外部記憶装置によれば、RAID5アレ
15 ID5ディスクを記憶ディスク部とする場合、送出キュー中に複数の書き込みコマンドがあれば、それらに伴う上記パリティ更新における読み出し処理が、各ディスク装置に分散し処理時間が均等化するように、上記第1の方法、第2の方法を選択する。これにより、書き込みコマンドに伴う読み出し処理が各ディスク装置間において並列に処理され、コマンド処理の効率が向上する
20 という効果が得られる。

【0073】本発明の外部記憶装置によれば、RAID3アレ
25 ID3ディスクを記憶ディスク部とする場合、データの分散度合いが高まり、コマンド処理が各ディスク装置間において並列に処理される可能性が高くなり、コマンド処理の効率が向上するという効果が得られる。

【0074】本発明の外部記憶装置によれば、シーク時間及び回転時間の実測値を利用することにより、コマ
30 ンド処理時間予測値の精度が高まり、外部記憶制御部によるコマンド処理の効率向上のための制御精度が上がる
40

いう効果がある。

【図面の簡単な説明】

【図1】 本発明が適用される共有外部記憶装置の模式的な構成図。

【図2】 共有外部記憶装置の内部処理を説明するためのブロック図。

【図3】 コマンドキューイング部の処理を説明するための送出キュー内のコマンド列の模式図。

【図4】 コマンドキューイング処理を説明するための送出キュー内のコマンド列の1例を示す模式図。

【図5】 コマンドキューイング処理を説明するタイミングチャート。

【図6】 電源オン時における自己診断の動作を説明するフローチャート。

【図7】 実際のコマンド発行処理におけるコマンドキューイング部の動作を説明するフローチャート。

【図8】 タイムアウト処理を説明する模式図。

【図9】 本発明が適用される他の共有外部記憶装置の模式的な構成図。

【図10】 RAID1ディスク・アレ
20 ID1ディスク・アレ
25 ID1ディスク・アレ
30 ID1ディスク・アレ
35 ID1ディスク・アレ
40 ID1ディスク・アレ
45 ID1ディスク・アレ
50 ID1ディスク・アレ
55 ID1ディスク・アレ
60 ID1ディスク・アレ
65 ID1ディスク・アレ
70 ID1ディスク・アレ
75 ID1ディスク・アレ
80 ID1ディスク・アレ
85 ID1ディスク・アレ
90 ID1ディスク・アレ
95 ID1ディスク・アレ
100 ID1ディスク・アレ
105 ID1ディスク・アレ
110 ID1ディスク・アレ
115 ID1ディスク・アレ
120 ID1ディスク・アレ
125 ID1ディスク・アレ
130 ID1ディスク・アレ
135 ID1ディスク・アレ
140 ID1ディスク・アレ
145 ID1ディスク・アレ
150 ID1ディスク・アレ
155 ID1ディスク・アレ
160 ID1ディスク・アレ
165 ID1ディスク・アレ
170 ID1ディスク・アレ
175 ID1ディスク・アレ
180 ID1ディスク・アレ
185 ID1ディスク・アレ
190 ID1ディスク・アレ
195 ID1ディスク・アレ
200 ID1ディスク・アレ
205 ID1ディスク・アレ
210 ID1ディスク・アレ
215 ID1ディスク・アレ
220 ID1ディスク・アレ
225 ID1ディスク・アレ
230 ID1ディスク・アレ
235 ID1ディスク・アレ
240 ID1ディスク・アレ
245 ID1ディスク・アレ
250 ID1ディスク・アレ
255 ID1ディスク・アレ
260 ID1ディスク・アレ
265 ID1ディスク・アレ
270 ID1ディスク・アレ
275 ID1ディスク・アレ
280 ID1ディスク・アレ
285 ID1ディスク・アレ
290 ID1ディスク・アレ
295 ID1ディスク・アレ
300 ID1ディスク・アレ
305 ID1ディスク・アレ
310 ID1ディスク・アレ
315 ID1ディスク・アレ
320 ID1ディスク・アレ
325 ID1ディスク・アレ
330 ID1ディスク・アレ
335 ID1ディスク・アレ
340 ID1ディスク・アレ
345 ID1ディスク・アレ
350 ID1ディスク・アレ
355 ID1ディスク・アレ
360 ID1ディスク・アレ
365 ID1ディスク・アレ
370 ID1ディスク・アレ
375 ID1ディスク・アレ
380 ID1ディスク・アレ
385 ID1ディスク・アレ
390 ID1ディスク・アレ
395 ID1ディスク・アレ
400 ID1ディスク・アレ
405 ID1ディスク・アレ
410 ID1ディスク・アレ
415 ID1ディスク・アレ
420 ID1ディスク・アレ
425 ID1ディスク・アレ
430 ID1ディスク・アレ
435 ID1ディスク・アレ
440 ID1ディスク・アレ
445 ID1ディスク・アレ
450 ID1ディスク・アレ
455 ID1ディスク・アレ
460 ID1ディスク・アレ
465 ID1ディスク・アレ
470 ID1ディスク・アレ
475 ID1ディスク・アレ
480 ID1ディスク・アレ
485 ID1ディスク・アレ
490 ID1ディスク・アレ
495 ID1ディスク・アレ
500 ID1ディスク・アレ
505 ID1ディスク・アレ
510 ID1ディスク・アレ
515 ID1ディスク・アレ
520 ID1ディスク・アレ
525 ID1ディスク・アレ
530 ID1ディスク・アレ
535 ID1ディスク・アレ
540 ID1ディスク・アレ
545 ID1ディスク・アレ
550 ID1ディスク・アレ
555 ID1ディスク・アレ
560 ID1ディスク・アレ
565 ID1ディスク・アレ
570 ID1ディスク・アレ
575 ID1ディスク・アレ
580 ID1ディスク・アレ
585 ID1ディスク・アレ
590 ID1ディスク・アレ
595 ID1ディスク・アレ
600 ID1ディスク・アレ
605 ID1ディスク・アレ
610 ID1ディスク・アレ
615 ID1ディスク・アレ
620 ID1ディスク・アレ
625 ID1ディスク・アレ
630 ID1ディスク・アレ
635 ID1ディスク・アレ
640 ID1ディスク・アレ
645 ID1ディスク・アレ
650 ID1ディスク・アレ
655 ID1ディスク・アレ
660 ID1ディスク・アレ
665 ID1ディスク・アレ
670 ID1ディスク・アレ
675 ID1ディスク・アレ
680 ID1ディスク・アレ
685 ID1ディスク・アレ
690 ID1ディスク・アレ
695 ID1ディスク・アレ
700 ID1ディスク・アレ
705 ID1ディスク・アレ
710 ID1ディスク・アレ
715 ID1ディスク・アレ
720 ID1ディスク・アレ
725 ID1ディスク・アレ
730 ID1ディスク・アレ
735 ID1ディスク・アレ
740 ID1ディスク・アレ
745 ID1ディスク・アレ
750 ID1ディスク・アレ
755 ID1ディスク・アレ
760 ID1ディスク・アレ
765 ID1ディスク・アレ
770 ID1ディスク・アレ
775 ID1ディスク・アレ
780 ID1ディスク・アレ
785 ID1ディスク・アレ
790 ID1ディスク・アレ
795 ID1ディスク・アレ
800 ID1ディスク・アレ
805 ID1ディスク・アレ
810 ID1ディスク・アレ
815 ID1ディスク・アレ
820 ID1ディスク・アレ
825 ID1ディスク・アレ
830 ID1ディスク・アレ
835 ID1ディスク・アレ
840 ID1ディスク・アレ
845 ID1ディスク・アレ
850 ID1ディスク・アレ
855 ID1ディスク・アレ
860 ID1ディスク・アレ
865 ID1ディスク・アレ
870 ID1ディスク・アレ
875 ID1ディスク・アレ
880 ID1ディスク・アレ
885 ID1ディスク・アレ
890 ID1ディスク・アレ
895 ID1ディスク・アレ
900 ID1ディスク・アレ
905 ID1ディスク・アレ
910 ID1ディスク・アレ
915 ID1ディスク・アレ
920 ID1ディスク・アレ
925 ID1ディスク・アレ
930 ID1ディスク・アレ
935 ID1ディスク・アレ
940 ID1ディスク・アレ
945 ID1ディスク・アレ
950 ID1ディスク・アレ
955 ID1ディスク・アレ
960 ID1ディスク・アレ
965 ID1ディスク・アレ
970 ID1ディスク・アレ
975 ID1ディスク・アレ
980 ID1ディスク・アレ
985 ID1ディスク・アレ
990 ID1ディスク・アレ
995 ID1ディスク・アレ
1000 ID1ディスク・アレ

【図11】 RAID5ディスク・アレ
20 ID5ディスク・アレ
25 ID5ディスク・アレ
30 ID5ディスク・アレ
35 ID5ディスク・アレ
40 ID5ディスク・アレ
45 ID5ディスク・アレ
50 ID5ディスク・アレ
55 ID5ディスク・アレ
60 ID5ディスク・アレ
65 ID5ディスク・アレ
70 ID5ディスク・アレ
75 ID5ディスク・アレ
80 ID5ディスク・アレ
85 ID5ディスク・アレ
90 ID5ディスク・アレ
95 ID5ディスク・アレ
100 ID5ディスク・アレ
105 ID5ディスク・アレ
110 ID5ディスク・アレ
115 ID5ディスク・アレ
120 ID5ディスク・アレ
125 ID5ディスク・アレ
130 ID5ディスク・アレ
135 ID5ディスク・アレ
140 ID5ディスク・アレ
145 ID5ディスク・アレ
150 ID5ディスク・アレ
155 ID5ディスク・アレ
160 ID5ディスク・アレ
165 ID5ディスク・アレ
170 ID5ディスク・アレ
175 ID5ディスク・アレ
180 ID5ディスク・アレ
185 ID5ディスク・アレ
190 ID5ディスク・アレ
195 ID5ディスク・アレ
200 ID5ディスク・アレ
205 ID5ディスク・アレ
210 ID5ディスク・アレ
215 ID5ディスク・アレ
220 ID5ディスク・アレ
225 ID5ディスク・アレ
230 ID5ディスク・アレ
235 ID5ディスク・アレ
240 ID5ディスク・アレ
245 ID5ディスク・アレ
250 ID5ディスク・アレ
255 ID5ディスク・アレ
260 ID5ディスク・アレ
265 ID5ディスク・アレ
270 ID5ディスク・アレ
275 ID5ディスク・アレ
280 ID5ディスク・アレ
285 ID5ディスク・アレ
290 ID5ディスク・アレ
295 ID5ディスク・アレ
300 ID5ディスク・アレ
305 ID5ディスク・アレ
310 ID5ディスク・アレ
315 ID5ディスク・アレ
320 ID5ディスク・アレ
325 ID5ディスク・アレ
330 ID5ディスク・アレ
335 ID5ディスク・アレ
340 ID5ディスク・アレ
345 ID5ディスク・アレ
350 ID5ディスク・アレ
355 ID5ディスク・アレ
360 ID5ディスク・アレ
365 ID5ディスク・アレ
370 ID5ディスク・アレ
375 ID5ディスク・アレ
380 ID5ディスク・アレ
385 ID5ディスク・アレ
390 ID5ディスク・アレ
395 ID5ディスク・アレ
400 ID5ディスク・アレ
405 ID5ディスク・アレ
410 ID5ディスク・アレ
415 ID5ディスク・アレ
420 ID5ディスク・アレ
425 ID5ディスク・アレ
430 ID5ディスク・アレ
435 ID5ディスク・アレ
440 ID5ディスク・アレ
445 ID5ディスク・アレ
450 ID5ディスク・アレ
455 ID5ディスク・アレ
460 ID5ディスク・アレ
465 ID5ディスク・アレ
470 ID5ディスク・アレ
475 ID5ディスク・アレ
480 ID5ディスク・アレ
485 ID5ディスク・アレ
490 ID5ディスク・アレ
495 ID5ディスク・アレ
500 ID5ディスク・アレ
505 ID5ディスク・アレ
510 ID5ディスク・アレ
515 ID5ディスク・アレ
520 ID5ディスク・アレ
525 ID5ディスク・アレ
530 ID5ディスク・アレ
535 ID5ディスク・アレ
540 ID5ディスク・アレ
545 ID5ディスク・アレ
550 ID5ディスク・アレ
555 ID5ディスク・アレ
560 ID5ディスク・アレ
565 ID5ディスク・アレ
570 ID5ディスク・アレ
575 ID5ディスク・アレ
580 ID5ディスク・アレ
585 ID5ディスク・アレ
590 ID5ディスク・アレ
595 ID5ディスク・アレ
600 ID5ディスク・アレ
605 ID5ディスク・アレ
610 ID5ディスク・アレ
615 ID5ディスク・アレ
620 ID5ディスク・アレ
625 ID5ディスク・アレ
630 ID5ディスク・アレ
635 ID5ディスク・アレ
640 ID5ディスク・アレ
645 ID5ディスク・アレ
650 ID5ディスク・アレ
655 ID5ディスク・アレ
660 ID5ディスク・アレ
665 ID5ディスク・アレ
670 ID5ディスク・アレ
675 ID5ディスク・アレ
680 ID5ディスク・アレ
685 ID5ディスク・アレ
690 ID5ディスク・アレ
695 ID5ディスク・アレ
700 ID5ディスク・アレ
705 ID5ディスク・アレ
710 ID5ディスク・アレ
715 ID5ディスク・アレ
720 ID5ディスク・アレ
725 ID5ディスク・アレ
730 ID5ディスク・アレ
735 ID5ディスク・アレ
740 ID5ディスク・アレ
745 ID5ディスク・アレ
750 ID5ディスク・アレ
755 ID5ディスク・アレ
760 ID5ディスク・アレ
765 ID5ディスク・アレ
770 ID5ディスク・アレ
775 ID5ディスク・アレ
780 ID5ディスク・アレ
785 ID5ディスク・アレ
790 ID5ディスク・アレ
795 ID5ディスク・アレ
800 ID5ディスク・アレ
805 ID5ディスク・アレ
810 ID5ディスク・アレ
815 ID5ディスク・アレ
820 ID5ディスク・アレ
825 ID5ディスク・アレ
830 ID5ディスク・アレ
835 ID5ディスク・アレ
840 ID5ディスク・アレ
845 ID5ディスク・アレ
850 ID5ディスク・アレ
855 ID5ディスク・アレ
860 ID5ディスク・アレ
865 ID5ディスク・アレ
870 ID5ディスク・アレ
875 ID5ディスク・アレ
880 ID5ディスク・アレ
885 ID5ディスク・アレ
890 ID5ディスク・アレ
895 ID5ディスク・アレ
900 ID5ディスク・アレ
905 ID5ディスク・アレ
910 ID5ディスク・アレ
915 ID5ディスク・アレ
920 ID5ディスク・アレ
925 ID5ディスク・アレ
930 ID5ディスク・アレ
935 ID5ディスク・アレ
940 ID5ディスク・アレ
945 ID5ディスク・アレ
950 ID5ディスク・アレ
955 ID5ディスク・アレ
960 ID5ディスク・アレ
965 ID5ディスク・アレ
970 ID5ディスク・アレ
975 ID5ディスク・アレ
980 ID5ディスク・アレ
985 ID5ディスク・アレ
990 ID5ディスク・アレ
995 ID5ディスク・アレ
1000 ID5ディスク・アレ

【図12】 RAID5ディスク・アレのセクタ・アドレスの割付けマップを表す模式図。

【図13】 従来のRAID3ディスク・アレのセクタ・アドレスの割付けマップを表す模式図。

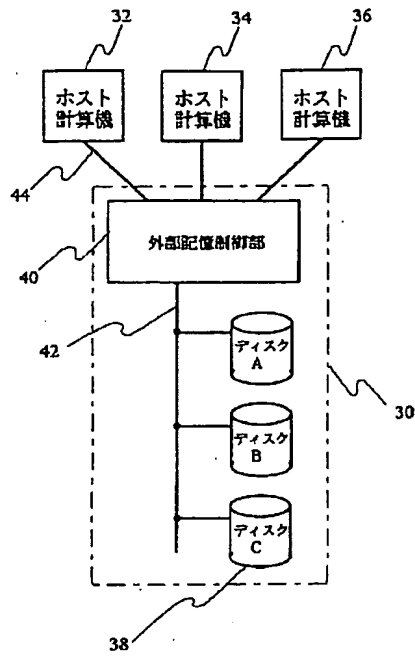
【図14】 実施形態におけるRAID3ディスク・アレのセクタ・アドレスの割付けマップを表す模式図。

【図15】 従来の外部記憶装置の模式的な構成図。

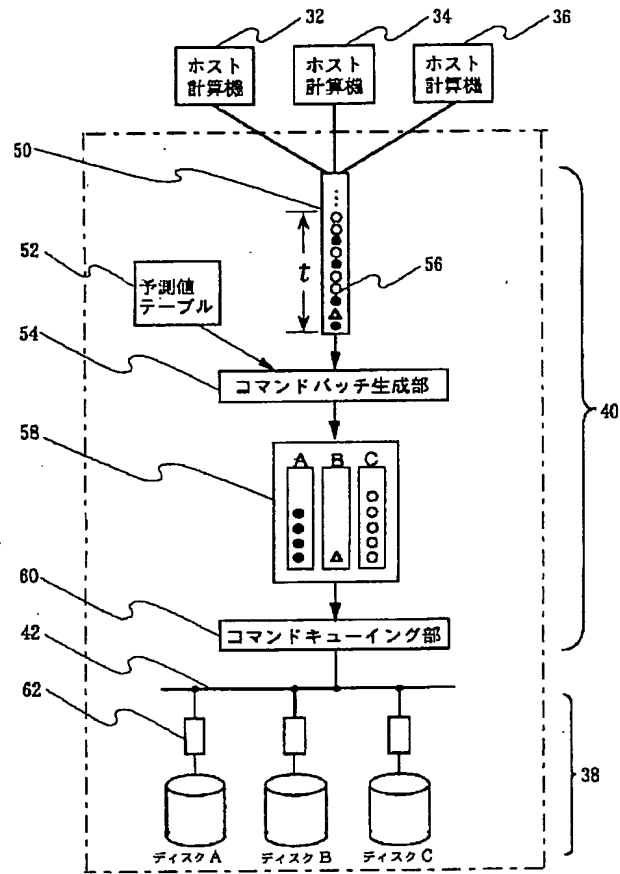
【符号の説明】

30 共有外部記憶装置、32、34、36 ホスト計算機、38 ディスク装置、40 外部記憶制御部、42、44 SCSIバス、50 受入キュー、52 予測値テーブル、54 コマンドバッチ生成部、58 送出キュー、60 コマンドキューイング部、62 ディスク側キュー、250、252 外部記憶制御部、256バス、358、360 RAID1ディスク・アレ、458、460 RAID5ディスク・アレ。

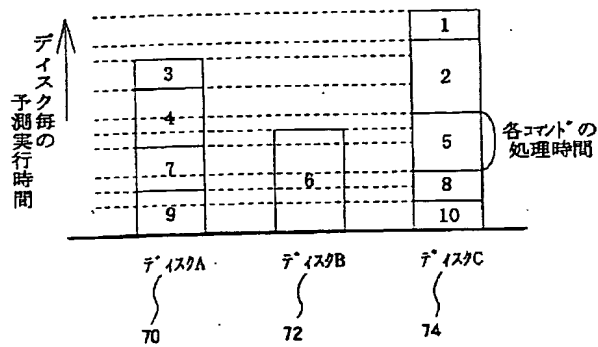
【図1】



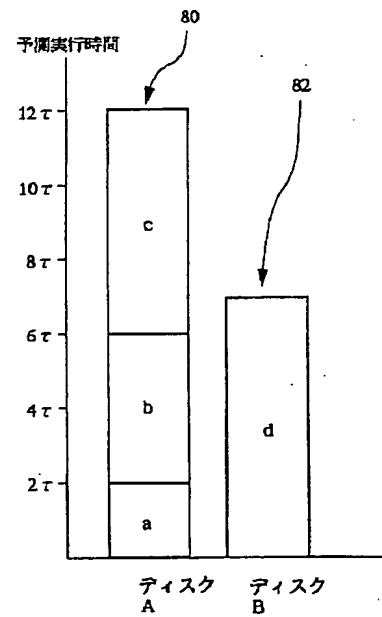
【図2】



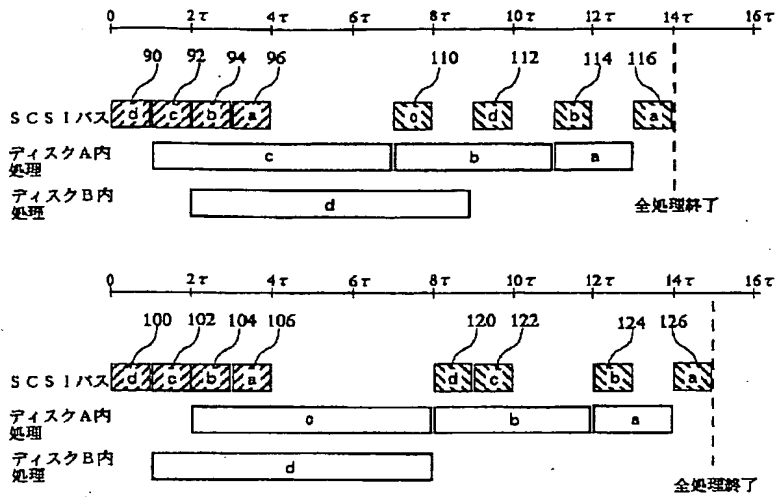
【図3】



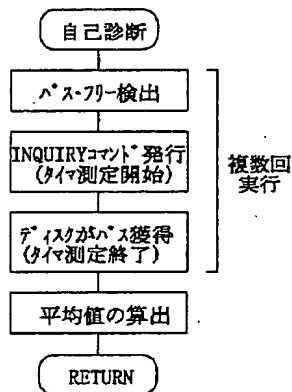
【図4】



【図5】



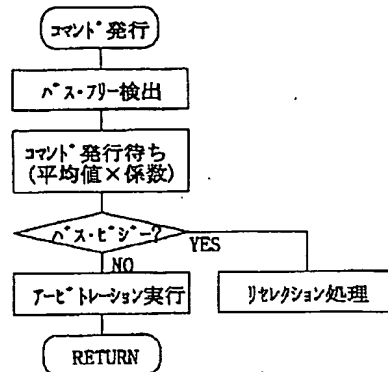
【図6】



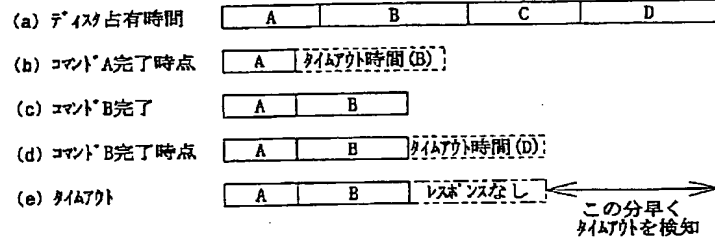
【図12】

DISK A	DISK B	DISK C	DISK D	DISK E
0	1	2	3	P
P	4	5	6	7
11	P	8	9	10
14	15	P	12	13
17	18	19	P	16
20	21	22	23	P
.
320	321	322	323	P
P	324	325	326	327
.
.
.

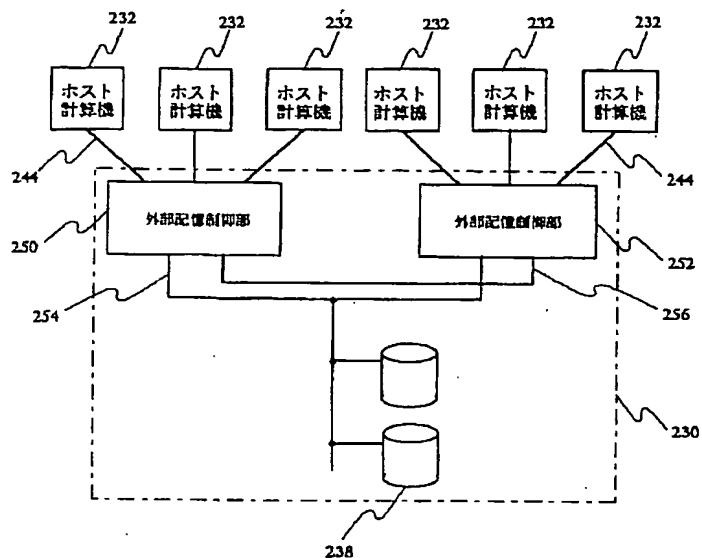
【図7】



【図8】



【図9】



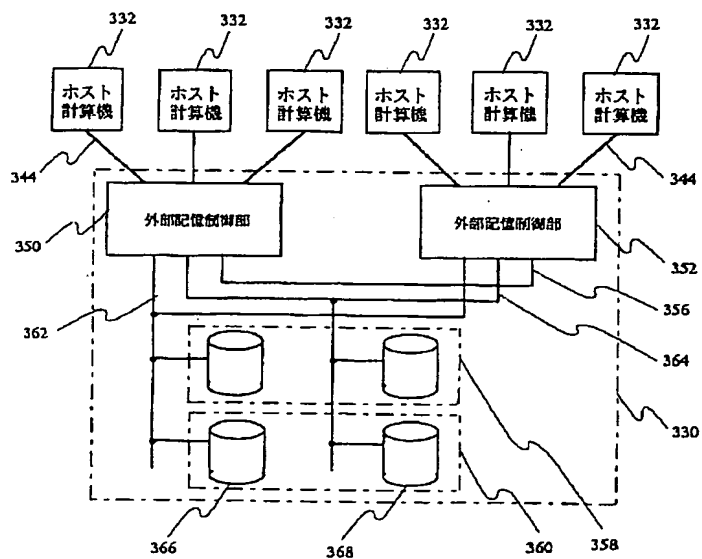
【図13】

DISK A	DISK B	DISK C	DISK D	DISK E
0	1	2	3	P
.
508	509	510	511	P
512	513	514	515	P
.
1020	1021	1022	1023	P
1024	1025	1026	1027	P
.
1532	1533	1534	1535	P
1536	1537	1538	1539	P
.
2044	2045	2046	2047	P
2048	2049	2050	2051	P
.
2556	2557	2558	2559	P
2560	2561	2562	2563	P
.
.

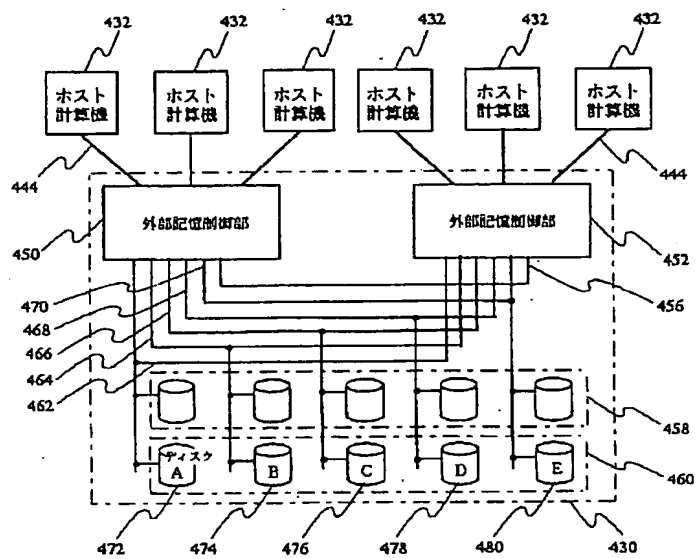
【図14】

DISK A	DISK B	DISK C	DISK D	DISK E
0	1	2	3	P
.
508	509	510	511	P
P	512	513	514	515
.
P	1020	1021	1022	1023
1027	P	1024	1025	1026
.
1535	P	1532	1533	1534
1538	1539	P	1536	1537
.
2048	2047	P	2044	2045
2049	2050	2051	P	2048
.
2557	2558	2559	P	2556
2560	2561	2562	2563	P
.
.

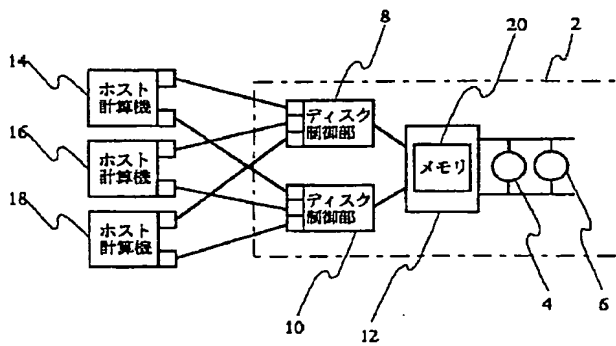
【図10】



【図11】



【図15】



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☐ BLACK BORDERS

☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

☒ FADED TEXT OR DRAWING

☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING

☐ SKEWED/SLANTED IMAGES

☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS

☐ GRAY SCALE DOCUMENTS

☒ LINES OR MARKS ON ORIGINAL DOCUMENT

☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.